Stochastic Models in Probability and Statistics
Vol. 2, No. 2, 2025, pp. 221–231. Research Paper

Probabilistic metric structures on birth-death population models

T. K. Murali[†*], P. K. Pramod[‡], P. B. Vinod Kumar[§]

[†]*Department of Mathematics, KKTM Govt. College, Pullut, Kerala, India*
[†]*Department of Mathematics, Maharaja's College, Ernakulam, Kerala, India*
[‡]*Department of Mathematics, Panambilly Memorial Govt. College, Chalakkudy, Kerala, India*
[‡]*Department of Mathematics, Maharaja's College, Ernakulam, Kerala, India*
[§]*Department of Basic Science and Humanities, Muthoot Institute of Technology and Science(MITS), Kochi, Kerala, India*

Email(s): muralitk.maths@gmail.com, pramodnair.2007@gmail.com,vinodkumarpb@mgits.ac.in

Abstract. In this paper, we introduce a probabilistic metric space structure for a population genetics model that incorporates mutation effects. Several properties of this structure are investigated in relation to the mutation process.

Keywords: Chapman-Kolmogorov equation, Markov Chains, Population genetics, Probabilistic metric spaces.

1 Introduction

A metric space is a nonempty set in which a non-negative value is assigned as the distance between each pair of points. In a probabilistic metric space, instead of assigning a deterministic non-negative number, a distance distribution function is associated with each pair of points. This function can sometimes be interpreted as the probability that the distance between the points does not exceed a given value.
Birth-death processes are stationary Markov processes in which only two types of transitions are possible, namely birth and death. It is typically assumed that, within a sufficiently small time interval, at most one birth or one death can occur. Using the probabilities of a single birth or death, transition probabilities are defined and a transition probability matrix is constructed. When the state space of a birth-death process is a metric space, a probabilistic metric space structure can be induced by combining the underlying metric with the transition probabilities.

Examining changes in the genetic composition of a population as a result of various evolutionary processes, including selection, mutation, migration, and random genetic drift, is one of the central objectives of theoretical population genetics. Consider a population of haploid individuals with a constant fixed size N. We focus on a single locus that can carry two types of alleles, A and B. Different alleles may lead to differences in fitness, meaning that some individuals are more strongly selected than others. If mutations can occur in both directions, the population will always remain polymorphic, and in this case

*Corresponding author
Received: 17 June 2025 / Accepted: 20 December 2025
DOI: 10.22067/smps.2025.94103.1048

the stationary distribution becomes the primary object of interest. To study these phenomena statistically, a suitable mathematical model is required. When attention is restricted to individuals of type A, mutation from A to B may be interpreted as a death event, while mutation from B to A corresponds to a birth event. Birth and death rates can be defined mathematically, and the mutation process can thus be modeled as a birth-death process. Using the transition probabilities associated with this process, the population consisting of two types of individuals can be endowed with a probabilistic metric space structure, allowing further properties of the mutation dynamics to be explored from this perspective. This construction is carried out in three different ways with slight modifications. In the first case, only the possible population sizes at different time points are considered as distinct states of the process. In the second and third cases, individual elements are treated as distinct, and the power set of the population is taken as the state space. In each case, a probabilistic metric space is constructed by taking into account the corresponding variations.

In Schweizer and Sklar (2005), probabilistic metric spaces are introduced and studied in detail. Markov chains and their fundamental properties are discussed in Bhat and Miller (1984), while applications of Markov chains in genetics are presented in Karlin (1975). The probabilistic metric space structure induced by stationary Markov processes is investigated in Marcus (1977). A similar idea is developed in Moynihan (1976), where stationary Markov processes with metric state spaces are considered. In Murali et al. (2022), three different cases of probabilistic metric spaces induced by birth-death processes are examined, along with related results.

In the present paper, we discuss simple examples of probabilistic metric spaces arising from genetic models and explore the applicability of the results in Murali et al. (2022) to the modeling of mutation processes in population genetics.

2 Preliminaries

2.1 Markov chains

In this paper, we employ Markov chains and their transition probabilities as follows. The number of members in the population at each stage is taken as the state space. Thus, the state space is $S = \{1, 2, \ldots\}$. Let $\{X_n\}_{n \in \mathbb{N} \cup \{0\}}$ denote a Markov chain defined on this state space. The probability distribution of X_{n+1} depends only on the value of X_n at time n and not on the previous states. Formally,

$$P\big(X_{n+1} = a_{n+1} \mid X_0 = a_0, X_1 = a_1, \ldots, X_n = a_n\big) = P\big(X_{n+1} = a_{n+1} \mid X_n = a_n\big).$$

We consider only time-homogeneous Markov chains. This property is expressed mathematically as

$$P(X_{n+1} = a_j \mid X_n = a_i) = P(X_{m+1} = a_j \mid X_m = a_i) \quad \text{for all } n, m.$$

Let $P_{ij}^{(n)} = P(X_{m+n} = a_j \mid X_m = a_i)$ denote the n-step transition probabilities. By the Chapman-Kolmogorov equations, we have $P_{ij}^{(r+s)} = \sum_{k \in S} P_{ik}^{(r)} P_{kj}^{(s)}$.

2.2 Metric spaces to probabilistic metric spaces

Definition 1. *A* metric space *is an ordered pair (S, d), where S is a nonempty set and $d \colon S \times S \to \mathbb{R}^+ \cup \{0\}$ is a function satisfying the following properties for all $a, b, c \in S$:*

(i) $d(a, b) \geq 0$;

(ii) $d(a, b) = 0$ *if and only if* $a = b$;

(iii) $d(a, b) = d(b, a)$;

(iv) $d(a,c) \le d(a,b) + d(b,c)$.

The value $d(a,b)$ is interpreted as the distance between the points a and b. Thus, in a metric space, a deterministic nonnegative distance is assigned to each pair of points. In contrast, in a probabilistic metric space, a distance distribution function is assigned to each pair of points. This function may be interpreted as describing the probability that the distance between the two points does not exceed a given value.

Definition 2. *A non decreasing function F defined on R^+ which satisfies $F(0) = 0$, $F(\infty) = 1$ and left continuous on $(0,\infty)$ is a distance distribution function.*

The collection of all distance distribution functions is denoted by Δ^+. For $a \ge 0$, the function $\varepsilon_a(x)$ is defined by

$$\varepsilon_a(x) = \begin{cases} 1, & x > a, \\ 0, & x \le a. \end{cases} \tag{1}$$

Definition 3. *A binary operation on Δ^+ is called a* triangle function *if it is commutative, associative, non-decreasing in each argument, and has ε_0 as the identity element.*

Equivalently, a map $\tau\colon \Delta^+ \times \Delta^+ \to \Delta^+$ is called a triangle function if it satisfies the following conditions:

1. *τ is commutative, that is, $\tau(F,G) = \tau(G,F)$, $\quad \forall F,G \in \Delta^+$;*
2. *τ is associative, that is, $\tau\big(\tau(F,G),H\big) = \tau\big(F,\tau(G,H)\big)$, $\quad \forall F,G,H \in \Delta^+$;*
3. *τ is non-decreasing in each argument, i.e., $F_1 \le F_2 \Rightarrow \tau(F_1,G) \le \tau(F_2,G)$, $\quad \forall G \in \Delta^+$, and $G_1 \le G_2 \Rightarrow \tau(F,G_1) \le \tau(F,G_2)$, $\quad \forall F \in \Delta^+$;*
4. *ε_0 is the identity element, that is, $\tau(F,\varepsilon_0) = F$, $\quad \forall F \in \Delta^+$.*

The product operation is a triangle function, defined by $(\tau(F,G))(x) = F(x)\,G(x)$. Also, the triangle function generated by the Łukasiewicz t-norm $T_m(a,b) = \max\{a+b-1, 0\}$ is denoted by τ_{T_m} and is defined as

$$(\tau_{T_m}(F,G))(x) = \sup_{0 \le u \le x} T_m\big(F(u), G(x-u)\big).$$

Definition 4. *Let $(S,\mathscr{F},\tau)$ be a triple, where S is a nonempty set, $\mathscr{F}\colon S \times S \to \Delta^+$ is a function, and τ is a triangle function. The triple $(S,\mathscr{F},\tau)$ is called a* probabilistic metric space *if, for all $p,q,r \in S$, the following conditions hold:*

(i) $\mathscr{F}(p,p) = \varepsilon_0$;

(ii) $\mathscr{F}(p,q) \ne \varepsilon_0$ *whenever* $p \ne q$;

(iii) $\mathscr{F}(p,q) = \mathscr{F}(q,p)$;

(iv) $\mathscr{F}(p,r) \ge \tau\big(\mathscr{F}(p,q), \mathscr{F}(q,r)\big)$.

If only conditions (iii) and (iv) are satisfied, then $(S,\mathscr{F},\tau)$ is called a pseudo probabilistic metric space.

The above definition is well defined. Indeed, since $\tau\colon \Delta^+ \times \Delta^+ \to \Delta^+$, for any $p,q,r \in S$ we have $\tau\big(\mathscr{F}(p,q), \mathscr{F}(q,r)\big) \in \Delta^+$. Moreover, Δ^+ is a partially ordered set, where for $A,B \in \Delta^+$, $A \le B$ if and only if $A(x) \le B(x)$ for all $x \ge 0$. Consequently, the inequality $\mathscr{F}(p,r) \ge \tau\big(\mathscr{F}(p,q), \mathscr{F}(q,r)\big)$ is meaningful and mathematically consistent.

Example 1. (Metric spaces as probabilistic metric spaces) Let (S,d) be a metric space. Take $F(p,q) = \varepsilon_{d(p,q)}$ where ε_d is given in (1). Also put $\tau(\varepsilon_a, \varepsilon_b) = \varepsilon_{a+b}$ then for any $p,q,r \in S$,

$$\tau(F(p,q),F(q,r)) = \varepsilon_{d(p,q)+d(q,r)}.$$

Hence, every metric space yields a Probabilistic metric space.

Example 2. (Genetic model) Consider the state space, $S = \{A,a\}$(two alleles at a single locus). We model mutation as a continuous- time Markov chain (CTMC) with constant rates:

$$A \xrightarrow{\mu} a,\ a \xrightarrow{\nu} A,$$

where $\mu, \nu > 0$ are mutation rates (per unit time). Let $(X_t)_{t\geq 0}$ be this CTMC and P_x denote probability with $X_0 = x$. Then, for $x,y \in S$, define

$$F_{xy}(t) = P_x\big(\exists\ s \leq t \text{ with } X_s = y\big),$$

as hitting distribution function. We will compute $F_{Aa}(t)$ and $F_{aA}(t)$ explicitly. Starting at A, the chain jumps away from A after an exponential time with parameter μ (the only outgoing rate). Thus, the time $T_{A\to a}$ to first hit a (which is the first jump time) is $\text{Exp}(\mu)$. Therefore,

$$F_{Aa}(t) = P(T_{A\to a} \leq t) = 1 - e^{-\mu t}.$$

Similarly, starting at a, we have

$$F_{aA}(t) = 1 - e^{-\nu t}.$$

By Definition 2, it readily follows that F_{Aa} and F_{aA} belong to Δ^+.

Remark 1. *$F_{xy} \neq \varepsilon_0$ for $x \neq y$ (since $F_{xy}(t) > 0$ for $t > 0$), and $F_{xx} = 1$ for all $t \geq 0$ if we define "hit itself by time t" trivially. To match the condition $F(p,p) = \varepsilon_0$, we instead define $F_{xx}(t)$ as the degenerate distribution: the time to hit itself is 0 with probability 1. Formally, take $F_{xy} = \varepsilon_0(t) = \mathbf{1}_{t\geq 0}$.*

• **Triangle inequality (probabilistic version).** Pick three states $p,q,r \in S$. There are only two nontrivial distinct-state choices; we will show the triangle-type inequality

$$F_{pr}(t) \geq \int_0^t dF_{pq}(S)F_{qr}(t-S). \tag{2}$$

This inequality is the probabilistic analogue of the triangle inequality and is obtained from the Markov property by conditioning on the first hitting time of q.

To Prove, (2), Let $T_{p\to q} := \inf\{s \geq 0 : X_s = q\}$ be the first hitting time of q starting from p. Then

$$\{\text{hit } r \text{ by time } t\} \supseteq \{T_{p\to q} \leq t \text{ and then hit } r \text{ within } t - T_{p\to q}\}.$$

Therefore, we have

$$\begin{aligned} F_{pr}(t) &= P_p(\exists\ s \leq t : X_s = r) \\ &\geq P_p(T_{p\to q} \leq t, \exists\ u \leq t - T_{p\to q} : X_{T_{p\to q}+u} = r). \end{aligned}$$

By the strong Markov property, conditional on $T_{p\to q} = s$ and $X_s = q$, the future evolution $\{X_{s+u} : u \geq 0\}$ has the same law as the chain started at q. Hence

$$\begin{aligned} P_p(T_{p\to q} \leq t, \exists\ u \leq t - T_{p\to q} : X_{T_{p\to q}+u} = r) &= \int_0^t P_p(T_{p\to q} \in ds)\ P_q(\exists\ u \leq t-s : X_u = r) \\ &= \int_0^t dF_{pq}(s)F_{qr}(t-s). \end{aligned}$$

Interpreting (2) as a triangle function inequality

Define a triangle function τ on Δ^+ by

$$(\tau(F,G))(t) := \int_0^t dF(s)\,G(t-s), \qquad F,G \in \Delta^+.$$

Here, the integral is the Lebesgue–Stieltjes convolution of F and G. The map τ satisfies the following properties:

- **Commutativity:** convolution is commutative.
- **Associativity:** convolution is associative.
- **Monotonicity:** τ is nondecreasing in each argument.
- **Identity:** ε_0 acts as a neutral element for convolution, since it is a point mass at 0.

Hence, τ is indeed a triangle function. The inequality (2) can then be written as $F_{pr}(t) \geq (\tau(F_{pq},F_{qr}))(t)$, so the probabilistic triangle axiom holds for this model.

Example 3. Multi-locus genotype (hypercube). We consider a classical mutation model from population genetics.
State space. The state space consists of binary sequences of length L, $S = \{0,1\}^L$, where each sequence represents a genotype. The Hamming distance $d_H(x,y)$ counts the number of sites at which two genotypes differ.
Mutation model. Each site mutates independently according to a symmetric two-state continuous-time Markov chain: a flip $0 \leftrightarrow 1$ occurs at rate μ. Consequently, the global process is the product of L independent two-state CTMCs.
Exact-time transition probabilities. If the process starts at genotype x, the probability of being exactly at genotype y at time t is, by independence across sites,

$$P_x(X_t = y) = p_{\text{flip}}(t)^{d_H(x,y)}\, p_{\text{same}}(t)^{L-d_H(x,y)},$$

where for a single site with symmetric mutation rate μ,

$$p_{\text{same}}(t) = \frac{1+e^{-2\mu t}}{2}, \qquad \text{and} \quad p_{\text{flip}}(t) = \frac{1-e^{-2\mu t}}{2}.$$

This exact-time transition probability does not define a distance distribution in the sense of Δ^+, since it is not monotone in t (it typically increases initially and then decreases as the chain approaches stationarity). Nevertheless, it serves as the transition kernel from which hitting-by-time distributions can be constructed.

Lower bound via a chosen intermediate genotype z. For $L \geq 2$, a closed-form expression for $F_{xy}(t)$ is generally unavailable. Nevertheless, we can derive a useful lower bound and verify the triangle inequality as before. For any intermediate genotype z,

$$F_{xy}(t) \geq \int_0^t dF_{xz}(s)\, P_z(X_{t-s} = y).$$

The proof is identical to the single-locus case: condition on the first hitting time of z at time s, and then require the process to be at y during the remaining time interval of length $t-s$. This yields a lower bound, since the probability of being at y at some time after s is at least the probability of being at y exactly at time t.
Taking a supremum over z, or iterating the above bound by further integration, produces bounds that satisfy the convolution inequality. Consequently, choosing τ to be the (Lebesgue–Stieltjes) convolution again yields a valid triangle inequality.

Example 4. Take the single-locus case with $\mu = 0.1$(per unit time). Then $F_{Aa}(t) = 1 - e^{-0.1t}$. After $t = 10$ units, $F_{Aa}(10) = 1 - e^{1} \approx 0.6321$. So 63.2% chance to have mutated by time 10.
With three states in 2-state model we only have pairs, but we can combine chains in series and use convolution to compute bounds.
For a 4-locus genome ($L = 4$) with identical per-site rate $\mu = 0.1$, the probability to be exactly at a particular genotype y at time t if Hamming distance k from start is

$$P_x(X_t = y) = \left(\frac{1 - e^{-0.2t}}{2}\right)^k \left(\frac{1 + e^{-0.2t}}{2}\right)^{4-k}.$$

Here the single-site exponent is $2\mu = 0.2$ because two-state diagonalization gives $e^{-2\mu t}$.

Remark 2. *For CTMC mutation models, the hitting-by-time distribution $F_{ij}(t) = P_i(\exists\ s : X_s = j)$ belongs to Δ^+ (it is nondecreasing, left continuous, limits 0 and 1). (Shown explicitly for the 2-state case and argued generally using CTMC properties.)*

Using the strong Markov property and conditioning on first hitting times, the convolution inequality

$$F_{ik}(t) \geq \int_0^t dF_{ij}(s)\ F_jk(t-s)$$

holds for all states i, j, k. Therefore, choosing the triangle function τ to be the Stieltjes convolution, the probabilistic triangle axiom $F(i,k) \geq \tau(F(i,j), F(j,k))$ is satisfied.

2.3 Transition probability matrix of mutation processes

Let X_n denote the number of individuals of type A at time n. The state space may be finite or infinite. For each i, let $b_i > 0$ denote the probability that a single individual of type B mutates to type A, and let $d_i > 0$ denote the probability that a type A individual mutates to type B. Initially, we assume that at each time step exactly one birth and one death occur.

$$P_{ij} = p[X_{n+1} = j / X_n = i] = \begin{cases} b_i & j = i+1 \\ d_i & j = i-1 \\ 1 - b_i - d_i & j = i \\ 0 & j \neq i-1, i, i+1 \end{cases} \quad \text{for } i = 0, 1, 2, \ldots$$

For finite case $P_{N,N+1} = b_N = 0$ where N is the total size of the population.

The transition matrix is $\begin{bmatrix} b_0 & 1-b_0 & 0 & 0 & 0 & \ldots & 0 \\ d_1 & 1-b_1-d_1 & b_1 & 0 & 0 & \ldots & 0 \\ 0 & d_2 & 1-d_2-b_2 & b_2 & 0 & \ldots & 0 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \cdots & 0 \\ 0 & 0 & 0 & 0 & 0 & \cdots & 1-d_N \end{bmatrix}$.

Lemma 1. *The above matrix is a valid transition probability matrix. i.e, $\sum_j P_{ij} = 1$ for all i.*

Proof. For $i = 0$, $\sum_j P_{ij} = 1 - b_0 + b_0 = 1$. For $i = 1$, $\sum_j P_{ij} = b_1 + 1 - (b_1 + d_1) + d_1 = 1$ and so on. For a fixed i, $\sum_j P_{ij} = b_i + 1 - (b_i + d_i) + d_i = 1$ for all i. □

3 Probabilistic metric space in population genetics

3.1 Mutation metrics on type A individuals

Consider a population consisting of N individuals. Each individual is of gene type A or gene type B. The state space is given by the number of type A individuals at a given time. A mutation of an individual from type A to type B is interpreted as a death, and the probability of a death in state i is denoted by d_i. Similarly, a mutation of an individual from type B to type A is interpreted as a birth, and the probability of a birth in state i is denoted by b_i.

Definition 5. *Let S be the state space $\{0,1,2,...\}$. P is the transition matrix for S that defines a mutation process $X_n, n \in N \cup \{0\}$. When this process is time homogeneous, for given two states i and j in S, define*

$$F_{ij}(t) = P[\text{ exist a time } s < t \text{ such that } X_s = j | X_0 = i]. \tag{3}$$

Remark 3. *Let $f_{ij}^{(s)}$ denote the probability that $X_0 = i$, the first time we visit state j is exactly at time s. Then $F_{ij}(t) = \sum_{s=1}^{u} f_{ij}^{(s)}, u < t$.*

Theorem 1. *Let S, P, $(X_n)_{n\in\mathbb{N}}$, and $F_{ij}(t)$ be defined as in the previous definition. Define the mapping $\mathscr{F} : S \times S \to \Delta$, where Δ denotes the space of cumulative distribution functions, by $\mathscr{F}(i,j) = F_{ij}$. Then the following properties hold:*

(i) $F_{ij}(0) = 0$;

(ii) F_{ij} *is left-continuous;*

(iii) $\mathscr{F}(i,i) = \varepsilon_0$ *for all* $i \in S$;

(iv) $\mathscr{F}(i,j) \neq \varepsilon_0$ *for* $i \neq j$;

(v) $\mathscr{F}(p,r) \geq \mathscr{F}(p,q)\mathscr{F}(q,r)$ *for all* $p,q,r \in S$, *where* ε_0 *is given in* (1).

Proof. By (3), part (i) is obvious. For (ii), we have

$$F_{ij}(t-h) = P[\text{ exist a time } s < t-h \text{ such that } X_s = j | X_0 = i].$$

Then, $\lim_{h\to 0} F_{ij}(t-h) = P[\text{exist a time } \quad s < t : X_s = j | X_0 = i] = F_{ij}(t)$, i.e, F_{ij} is left continuous. For (iii), as $F_{ii}(t) = P[\text{exist a time } \quad s < t : X_s = i | X_0 = i]$ for $s = 0$, $X_s = i$ when $X_0 = i$. Therefore

$$F_{ii}(t) = \begin{cases} 1 & t > 0 \\ 0 & t \leq 0. \end{cases}$$

(iv) Let $i \neq j$, it needs a positive time s_0 to reach state j from the state i. For $0 \leq t \leq s_0 \quad F_{ij}(t) = 0$ and obviously $F_{ij}(t) \neq \varepsilon_0$.

(v) For the different states p,q and r and time points t_1 and t_2, we have

$$\begin{aligned}
F_{pr}(t_1+t_2) &= P[\text{there exist a time } \ s<t_1+t_2 : X_s = r | X_0 = p] \\
&\geq P[\text{there exist a time } i<t_1 : X_i = q \text{ for the first time and } X_s = r \text{ for some } i<s<t_1+t_2 | X_0 = p] \\
&= \sum_{i=1}^{[t_1]} P[X_i = q \ \text{ for the first time and } \ X_s = r \ \text{ for some } \ s \in [i, t_1+t_2) | X_0 = p] \\
&= \sum_{i=1}^{[t_1]} P[X_i = q \text{ for the first time} | X_0 = p] P[X_s = r \text{ for some } s \in [i,t_1+t_2) | X_0 = p \text{ and } X_i = q \\
&\qquad\qquad \text{for the first time}] \\
&= \sum_{i=1}^{[t_1]} f_{pq}^i P[X_s = r \quad \text{for some} \quad s \in [i,t_1+t_2) | X_i = q \quad \text{for the first time}] \\
&= \sum_{i=1}^{[t_1]} f_{pq}^i P[X_s = r \quad \text{for} \quad s \in [0,t_1+t_2-i) | X_0 = q] \\
&\geq \sum_{i=1}^{[t_1]} f_{pq}^i P[X_s = r \quad \text{for some} \quad s \in [0,t_2) | X_0 = q] \\
&= F_{pq}(t_1) F_{qr}(t_2).
\end{aligned}$$

It means $\mathscr{F}(p,r) \geq \mathscr{F}(p,q)\mathscr{F}(q,r)$. □

Theorem 2. *Let $S, P, (X_n)_{n\in\mathbb{N}}$ and $F_{ij}(t)$ be defined as above. Then $(S, \mathscr{F}, \tau)$ is a probabilistic metric space with a non-symmetric distance function where $(\tau(F,G))(x) = F(x)G(x)$, when $P(T_{ij} < \infty) = 1$ where $T_{ij} = \{min(n \geq 0 : X_n = j | X_0 = i)\}$.*

Proof. It follows from Definition 4 and Theorem 1 □

3.2 Mutation metrics on type A subsets in finite populations

In the above model, each state represents the number of type A individuals in the population. However, different groups of type A individuals with the same cardinality are represented by the same number in this case. In the current construction, each distinct group of type A individuals is considered as a separate state. Initially, we assume that at most one birth and one death occur at a time. When the composition of the entire population is known, it is possible to determine the probability of a single birth, P_b, and a single death, P_d, at any given time point.

Definition 6. *Let P_b and P_d denote the probabilities of a single birth and a single death, respectively. At any time, the population of type A individuals can be represented as a subset of the whole population S, where $|S|$ denotes the total population size. For two subsets of type A individuals, S_i and S_j, we define the transition probability from S_i to S_j as $P_{ij} = P(X_{n+1} = S_j \mid X_n = S_i)$.*

Lemma 2. *Take $a_1 : S_j \subset S_i$ and $|S_i \backslash S_j| = 1$, $a_2 : S_i \subset S_j$ and $|S_j \backslash S_i| = 1$, $a_3 : |s_i| \neq 0$ or $|P|$, $a_4 : |S_0| = 0$ or $|P|$ and $a_5 : S_i = S_j$. Then,*

$$P_{ij} = \begin{cases} P_d \frac{1}{|S_i|} & for \quad a_1 \quad and \quad a_3 \\ P_b \frac{1}{|P|-|S_i|} & for \quad a_2 \quad and \quad a_3 \\ \frac{1}{|P|} & for \quad (a_1 \quad or \quad a_2) \quad and \quad a_4 \\ 1-(P_d+P_b) & for \quad a_5 \quad and \quad a_3 \\ 0 & otherwise \end{cases}$$

leads to a transition probability matrix.

Proof. For a fixed i and $|S_i| \neq 0$ or $|P|$.

$$P_{ii} = 1 - (P_d + P_b)$$

and

$$\sum_{j \neq i} P_{ij} = |S_i| P_d \frac{1}{|S_i|} + (|P| - |S_i|) P_b \frac{1}{(|P| - |S_i|)}$$
$$= P_d + P_b$$

(First term for $|S_i \backslash S_j| = 1$ and second term for $|S_j \backslash S_i| = 1$ and 0 for other cases)

$$\sum_j P_{ij} = 1 - (P_d + P_b) + P_d + P_b = 1$$

Similarly we can prove for other cases. □

Definition 7. *Let $\mathscr{P}$ be the population and State space S is the power set of $\mathscr{P}$. Let P be the transition matrix for S which defines a time homogeneous Markov chain $(X_n)_{n \in \mathbb{N}}$. For any two given subsets S_i and S_j, we define*

$$F_{ij}^S(t) = P[\text{ there exist a time } s < t \text{ such that } X_s = S_j | X_0 = S_i]$$

Remark 4. *If f_{ij}^s denote the probability that with $X_0 = i$, the first time we visit j is exactly at time s, then*

$$F_{ij}^S(t) = \sum_{s=1}^{u} f_{ij}^s, u < t$$

Theorem 3. *Theorem 3 Let $S, P, (X_n)_{n \in \mathbb{N}}$ and $F_{ij}(t)$ be defined as in Definition 7, then $(S, \mathscr{F}, \tau)$ is a probability metric space with a non-symmetric distance function where $(\tau(F,G))(x) = F(x)G(x)$, when $P(T_{ij} < \infty) = 1$ where $T_{ij} = \{min(n \geq 0 : X_n = S_j | X_0 = S_i)\}$*

Proof. It follows from Theorem 2 and Definition 7, replacing i and j by S_i and S_j respectively. □

Theorem 4. *Let S_1 and S_2 be two states or subsets of $\mathscr{P}$. Let $t_a = |S_2 \backslash S_1|$ is the number of new births not in S_1 and $t_b = |S_1 \backslash S_2|$ be the number of deaths or the number of elements in S_1 but not in S_2. Let $t^* = t_a + t_b = |S_2 \backslash S_1| + |S_1 \backslash S_2|$. Then, $F_{12}(t) = 0$ for $t < t^*$. S_2 can be obtained from S_1 within a minimum of t^* steps.*

Proof. $t_a = |S_2 - S_1|$ and $t_b = |S_1 - S_2|$. Minimum number of steps or time to reach S_2 from $S_1 = t_a + t_b = t^*$ i.e, for $t < t^*$ it is impossible to reach S_2 from S_1 i.e, $F_{12}(t) = P[$ there exist a time $s < t^*$ such that $X_s = S_2 | X_0 = S_1] = 0$. □

3.3 Mutation metrics on a metric state space

In this case, we can consider the probabilistic metric space structure induced by the mutation process to probabilistically estimate the differences between states in the future due to changes in the initial state. Let $\mathscr{P}$ denote the population in a birth-death process, with the power set of $\mathscr{P}$ as the state space. A stationary Markov chain can then be constructed as in the previous section.

Definition 8. *Let $S_i, S_j \in S$ and $d(S_i, S_j) = |S_i \backslash S_j| + |S_j \backslash S_i|$. Then, (S, d) is a metric space. Now for each t, define $\mathscr{F}^{(t)} : S \times S \to \Delta^+$ as follows.*

$$\mathscr{F}^{(t)}(S_i, S_i)(x) = \varepsilon_0, \forall i$$

and

$$\mathscr{F}^{(t)}(S_i, S_j)(x) = P\{\text{distance between the states is less than } x \text{ at time } t, \text{ if the process starts at } S_i \text{ and } S_j\}.$$

Using the results of Moynihan (1976), we obtain the following conclusions.

Theorem 5. *For each t, $(S,\mathscr{F}^{(t)},\tau)$ is a pseudo probabilistic metric space under $\tau=\tau_{Tm}$.*

Proof. First, consider the case $S_i=S_j$. Then, $d(S_i,S_j)=0$ implies that

$$\text{Prob}\{\text{distance between states} < x \text{ for } x\geq 0 \text{ at time } t \text{ when starting at } S_i \text{ and } S_j\}=1 \text{ for } t\geq 0.$$

Hence, $\mathscr{F}^t(S_i,S_j)=\varepsilon_0$.
Next, consider the case $S_i\neq S_j$. Then $d(S_i,S_j)\neq 0$, but eventually the two states may reach the same state. Therefore, there exists a time t_0 such that the distance between the states can be zero, i.e.,

$$\mathscr{F}^{t_0}(S_i,S_j)=\varepsilon_0.$$

For fixed t, the left-continuity of $\mathscr{F}^t(S_i,S_j)$ follows from

$$\lim_{h\to 0}\mathscr{F}^t(S_i,S_j)(x-h)=\text{Prob}\{\text{distance between states} \leq x \text{ at time } t\}=\mathscr{F}^t(S_i,S_j)(x).$$

Symmetry is straightforward: for fixed t,

$$\mathscr{F}^t(S_i,S_j)(x)=\mathscr{F}^t(S_j,S_i)(x).$$

Finally, consider the triangle inequality for any $S_i,S_j,S_k\in S$ and $t\geq 0$:

$$\begin{aligned}\mathscr{F}^t(S_i,S_k)(x_1+x_2)&=\text{Prob}\{\text{distance between states at time } t<x_1+x_2 \text{ starting at } S_i \text{ and } S_k\}\\&\geq \text{Prob}\{\text{distance} < x_1 \text{ starting at } S_i,S_j \text{ and distance} < x_2 \text{ starting at } S_j,S_k\}\\&\geq \mathscr{F}^t(S_i,S_j)(x_1)+\mathscr{F}^t(S_j,S_k)(x_2)-1\\&\geq T_m\big(\mathscr{F}^t(S_i,S_j)(x_1),\mathscr{F}^t(S_j,S_k)(x_2)\big),\end{aligned}$$

where T_m denotes the Lukasiewicz t-norm. Hence, for $\tau=T_m$,

$$\mathscr{F}^t(S_i,S_k)\geq\tau\big(\mathscr{F}^t(S_i,S_j),\mathscr{F}^t(S_j,S_k)\big).$$

This verifies all conditions for $(S,\mathscr{F}^{(t)},\tau)$ to be a pseudo probabilistic metric space under $\tau=\tau_{Tm}$. □

Theorem 6. *Let $\mathscr{F}_{ij}^{(t)}=\mathscr{F}^t(S_i,S_j)(x)$. Then for any real x, any states S_i,S_j, and any $r,t\geq 0$, we have*

$$F_{ij}^{(r+t)}(x)=\sum_{m=0}^{\infty}\sum_{n=0}^{\infty}P_{im}(r)P_{jn}(r)F_{mn}^{(t)}(x),$$

or in matrix form,

$$F^{(r+t)}(x)=P(r)F^t(x)\big(P(r)\big)^T,$$

where $P(r)^T$ denotes the transpose of the matrix $P(r)=(P_{ij}(r))$.

Proof. By definition,

$$\mathscr{F}^t(S_i,S_j)(x)=F_{ij}^t(x)=\text{Prob}\{\text{distance between states} < x \text{ at time } t \text{ starting at } S_i \text{ and } S_j\}.$$

This can be expressed as

$$F_{ij}^t(x)=\sum_{k=0}^{\infty}P_{ik}(t)\sum_{\substack{l\\ d(l,k)<x}}P_{jl}(t).$$

For time $r+t$, we have

$$F_{ij}^{r+t}(x) = \sum_{k=0}^{\infty} P_{ik}(r+t) \sum_{\substack{l \\ d(l,k)<x}} P_{jl}(r+t). \tag{4}$$

Using the Chapman-Kolmogorov equation,

$$P_{ij}(r+t) = \sum_{k\in S} P_{ik}(r)P_{kj}(t),$$

we substitute into (4):

$$F_{ij}^{r+t}(x) = \sum_{k=0}^{\infty} \left[\sum_{m\in S} P_{im}(r)P_{mk}(t)\right] \sum_{\substack{l \\ d(l,k)<x}} \left[\sum_{n\in S} P_{jn}(r)P_{nl}(t)\right].$$

Interchanging the order of summation gives

$$\begin{aligned} F_{ij}^{r+t}(x) &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} P_{im}(r)P_{jn}(r) \left[\sum_{k=0}^{\infty} P_{mk}(t) \sum_{\substack{l \\ d(l,k)<x}} P_{nl}(t)\right] \\ &= \sum_{m=0}^{\infty}\sum_{n=0}^{\infty} P_{im}(r)P_{jn}(r)F_{mn}^{t}(x), \end{aligned}$$

as claimed.

□

Acknowledgement

We express our gratitude to the reviewers for their thorough reading of the manuscript and their perceptive remarks.

References

Marcus, P. S. (1977). Probabilistic metric spaces constructed from stationary Markov chains. *Aequationes Mathematicae*, **15**, 169–171.

Moynihan, R. (1976). Probabilistic metric spaces induced by Markov chains. *Zeitschrift für Wahrscheinlichkeitstheorie und Verwandte Gebiete*, **35**, 177–187.

Karlin, S. (1975). *A First Course in Stochastic Processes, Second edition.* Academic Press, New York.

Bhat, U. N., and Miller, G. K. (2002). *Elements of Applied Stochastic Processes* (Vol. 3). Wiley-Intersciencem, Hoboken, NJ.

Schweizer, B., and Sklar, A. (2005). *Probabilistic Metric Spaces*, Dover Publications, New York.

Murali, T. K., Vinod Kumar, P. B. and Pramod, P. K. (2022). Birth and death processes: a probabilistic metric space approach. Mathematics of Intelligent Computing and Data Science, Springer Proceedings in Mathematics& Statistics, Vol. 484, 223–235.